Forum Series on the Role of Institutions in Promoting Economic Growth

Comments by Peter Murrell on Omar Azfar and Clifford Zinnes's *"Learning from Doing: A Methodology for Self-Evaluating Projects"*[1]

Forum 5: NIE-Based Toolkits for USAID Applications

Session 1

14 February 2003
Washington, D.C.

Let me first of all review what I see as the main points of the paper. The authors begin by observing that there is an absence of information feedback from projects. They then discuss randomized clinical trials, suggesting that this is a method that we should emulate when trying to secure that informational feedback. The authors argue that if these methods can be successfully emulated, then the effectiveness of foreign aid projects will improve. They also make the very important point that the increased decentralization of many governments greatly enhances the opportunities to conduct this type of work. Additionally, the authors provide a series of applications of this method, all of which I find interesting and credible. The practical implementation of this method in the form of a training program in the Philippines really does show how one of these kinds of experiments would work.

Having summarized the main points, let me outline what I find to be convincing in the paper. The basic point is the necessity of finding out which projects or implementation methods work better, and the real virtue of using experiments to find this out. Their examples do show clearly how decentralization can help in this whole process.

Now let me explain what I did not find convincing in the paper. I find that they undersold the basic points by overemphasizing formalism. In fact, reading through the paper one feels that the authors' emphasis on a too-scientific approach blurs the basic lessons that the authors are trying to convey. The way to make this blurring look stark is to observe that at the beginning of the paper, the authors correctly highlight the fact that there is very little systematic evidence collected thus far on the impact of variations in the design and implementation of USAID's various activities. Immediately thereafter, though, they argue for the relevance of one of the most developed scientific methodologies in the world—randomized clinical trials in modern medicine—to address this gap. This is quite a large gap to fill, and we may have more success with more modest ambitions—using less scientific, but more pragmatic means to evaluate projects. A simple alternative would be to collect data simply on what people who implement the projects think about particular projects relative to other projects. To begin to collect data, all one needs is something to measure—opinions, preferences, something.

Let us now examine each of the authors' criteria for conducting a PREP-project.

[1] The previous title for this paper, referred to below, is "Success by Design: The Science of Self-Evaluating Projects."

1) The reform must be implemented according to a clear randomization protocol
    - That would a worthy, but probably unattainable goal. Most progress in economics proceeds by trying to understand the non-random elements.
    - Since randomization is so expensive and requires so much effort, it may be more effective to try to get a sense of the non-random elements before conducting any experiment.
2) There must be a large number of individuals in the placebo group and the treatment group
    - We're primarily worried about whether a reform works—we would be happy if we had a reform that worked 60-70% of the time. For example, suppose some reform was conducted and evaluated, but non-random elements obscured a clear evaluation 25% of the time. If we were then to implement this reform in 10 regions, our evaluation would predict the correct answer 95% of the time.
    - So we don't need a large number to be able to learn a lot.
3) The treatment group must respond to the reform as though the reform were broadly implemented.
    - This is, of course, an ideal. In practice, it is not necessary—only that the direction of reform must be the same as it would have been had the reform been broadly implemented. This latter criterion is not a strong one at all.
4) The placebo group must continue to act as it would if the reform were not implemented.
    - Again, this is too strong. We only need reason to believe that the effect of the reform would be stronger on the control group than it would on the placebo group.
5) There must be no spillover effects between the control and the treatment groups.
    - In fact the spillover effects simply must be weak enough so that the direct effects drown out any spillover effects.
6) The effects of the reform must become evident in a short time.
    - Why? Ten years have passed since the various privatization programs in Russia were initiated, and we can still return to study the effects and effectiveness of different kinds of reform.
7) Evaluation should occur both before the reform and again long enough after the reform for its impact to be apparent.
    - Again, this is an ideal. In the best cases you can do this. But some reforms don't require pre-reform measurements. In many cases, you can learn a lot simply by assuming that most people started off in the same situation.

To conclude, I want to suggest that to improve the effectiveness of foreign aid delivery, we do not need to use an approach as rigorous as natural science. Economics, surely, will need many centuries before it will be as good as natural science is now.

We need, instead, practical tools that combine the insights from economic theory with the insights that statistics has given us. With these sorts of tools, we are likely to be

able to improve the design and delivery of programs with a likelihood of, say, 70% that they will in fact make a positive difference.

With this comments in mind, I would like to reconsider the title of the paper: "Success by Design: The Science of Self-Evaluating Projects."

Rather than success, we should aim for improvement. Rather than design, we should emphasize experimentation. Rather than "the", we should reasonably hope for "some." Rather than science, we should allow ourselves many methods.

This would give us: "Improvements by Experimentation: Some Methods to Evaluate Projects."